

~~CONFIDENTIAL~~

MAR 18 1954

MEMORANDUM FOR THE RECORD -- Management Improvement Staff

SUBJECT: Machine Translation

A. General Information

1. The problem of designing and developing a machine which can automatically scan a document written in one language, translate that language into one or more other languages and print the translations, has been discussed and studied for some time by various individuals here and abroad. Until recently only limited progress has been made in overcoming the difficulties of developing feasible mechanical and linguistic techniques. While the requirements of both techniques are difficult to meet, it is believed that the linguistic aspects of the problem will require more time and effort.
2. The solution of the machine phases of the problem has been greatly accelerated by the recent technological advances in the field of electronic data processing machines and as a result, those individuals engaged in research in the mechanical translation field, have been encouraged to press for a solution to the entire problem.
3. The electronic data processing machines now in use were designed primarily to solve complex mathematical problems in comparatively short periods of time; mathematical problems which formerly required months and years of computation, are now computed by these machines in a matter of hours and days. In addition to high speed operation, they include two operating characteristics which are essential and provide a partial solution to the machine translation problem. These characteristics are:
 - a. The ability to receive and store (remember) large quantities of data pertinent to a specific problem. In the case of machine translation, the storage facility would be used to store a language "dictionary", grammatical rules, syntactical rules and programmed instructions.
 - b. The ability to completely and automatically execute programmed instructions for processing the stored data to solve a specific problem.

~~CONFIDENTIAL~~

Document No. 13
11-27-28
By: 35

CONFIDENTIAL

The above operating characteristics are inherent in various types and makes of electronic data processing machines. However, the input medium now used in all types of these machines is either a punched card or a magnetic tape. While it is possible for a person without knowledge of any foreign language to transcribe the information, this operation is comparatively slow and costly. It is believed that for an entirely successful translation machine, it will be necessary to provide a mechanism which can automatically scan the material to be translated and feed it into the translation machine. Some progress has already been made in the development of such a mechanism, commonly referred to as a "character sensing" machine. It is felt that as a translation machine is being designed and developed, emphasis should be placed on the development of the character sensing mechanism so that it will be available for use with the completed translation machine.

B. Progress Report - Georgetown University and IBM Corporation Research

1. A practical demonstration of the progress which has been made to date in utilizing a conventional electronic data processing machine as a "translation machine" and in the development of the necessary linguistic techniques for machine translation was held at IBM World Headquarters, New York, New York, on 7 January 1954, under the auspices of the Georgetown University Institute of Linguistics and Languages and the IBM Corporation. Their research effort and the results of the progress demonstration can be described briefly as follows:
 - a. Under the direction of Dr. Leon Dostert, the Georgetown University Institute of Linguistics and Languages, jointly with members of the science and research staff of the International Business Machines Corporation under the direction of Dr. Cuthbert Hurd, has been conducting "a preliminary research project for testing in a practical manner certain principles and techniques for electro mechanical language conversion more commonly referred to as mechanical translation". Progress reports on this research have been furnished to this Agency by Dr. Dostert. In his last report dated 3 June 1953, it was indicated that an actual demonstration of the progress which had been made would be held in the near future. An outline of the research and development work in linguistics and electronic machine techniques, which had to be completed before holding the demonstration in New York, was contained in the progress report of 3 June 1953.
 - b. By a letter dated 11 December 1953 to Dr. [REDACTED] SIC, Dr. Dostert invited representatives of the Agency to

25X1A9a

~~CONFIDENTIAL~~

attend a private demonstration to be held on 7 January 1954 at IBM World headquarters, New York, New York. The purpose of the demonstration was to perform "actual machine translation of a limited segment of Russian and English and to review the major steps of the work completed to date".

- c. The demonstration was held as scheduled and successfully demonstrated that machine translation of one language into another was feasible. The demonstration itself consisted briefly of the following:
- (1) A glossary of 250 Russian words had been developed. Each word carried with it one or two English equivalents and three "grammatical" codes.
 - (2) Six basic rules of syntax for instructing the machine on how to apply the grammatical rules for each word had been established.
 - (3) A program for instructing the machine on how to "look up" Russian words in the glossary and apply the syntactical rules for omission, insertion, inversion, and multiple choice of English words had been developed by the IBM engineers.
 - (4) The glossary, syntactical rules and program were introduced into an IBM 701 Electronic Data Processing Machine which stored all of this data in its electro-static memory. After this was done, the machine was ready to accept Russian sentences for translation.
 - (5) Several preconceived Russian sentences were used for demonstrating the machine operation.
 - (6) The input medium chosen for purpose of this demonstration was an IBM card. The preconceived Russian sentences were, therefore, transcribed in punched-card form by a key punch operator. The operator did not know the Russian language and this operation was simply a matter of "copying" the Russian language sentences.
 - (7) After the cards were punched, they were fed into the 701 machine and in a matter of a few seconds the machine "read" each Russian word, located the words in the glossary, applied the rules of syntax for choosing the proper English equivalent, and printed the English words in their proper sequence to produce an intelligible translation of the Russian sentences.

~~CONFIDENTIAL~~

2. The demonstration definitely proved that an Electronic Data Processing Machine such as the 701 could be programmed to perform language conversion and it was also evident that the basic linguistic techniques which had been developed were feasible. Since the stated purpose of this demonstration was to show the progress which had been made, to date, in solving the mechanical translation problem, it can be said that substantial and encouraging progress has been made in the development of basic machine and linguistic techniques.
3. The demonstration itself might have been still more effective had the audience been requested, after the initial demonstration was completed, to improvise some Russian sentences for translation using some of the 250 words available in the glossary rather than basing the entire demonstration on sentences prepared prior to the demonstration. However, this is a minor point and should not be interpreted to detract in any degree from a noteworthy demonstration. The use of previously prepared material for demonstrating procedures and machine operation is common practice. This method is followed so that all the steps in an operational procedure can be demonstrated in continuous sequence without distracting the witnesses by stopping to prepare demonstration material.

C. Conclusions

1. As Dr. Doertert has pointed out ~~that~~ "mechanical translation" is at the Kitty Hawk stage. A great deal of time and effort must be still expended in research and development work over an estimated period of five years before all the specifications of the "ultimate" machine can be met. But, as the basic principles of airplane construction were demonstrated at Kitty Hawk, the basic principles for developing mechanical translation were demonstrated in New York.
2. However, it is believed that substantial economic and operational benefits might be derived from the use of not only the ultimate machine, but from an "interim" machine which could be constructed as operating characteristics short of those needed for the ultimate machine, are developed and found feasible.
3. For example, linguistic and machine specifications for the ultimate machine might be to produce a "smooth" translation of both technical and non-technical Russian into English and from technical and non-technical English into Russian. The Agency's major need might be benefited with the translation of only scientific and technical Russian into a "rough" English translation. Such being the case, it might prove profitable to have a machine constructed when that stage of its development is reached so that

~~CONFIDENTIAL~~

immediate benefits could be realized. The construction of machine with limited operating characteristics should not, however, preclude the Agency's continued interest and support of the program to produce the "ultimate" translation machine.

- h. It is also believed that a "translation machine" incorporating the operating characteristics discussed in this memorandum, has other legitimate uses in addition to its primary purpose of language conversion. For example, it could be used as a coding and decoding machine in connection with the machine indexing of the informational content of documents. In this case, the "dictionary" to be stored in the machine would subject words and their searching machine code equivalents. The operations for automatic coding would be similar to those for language conversion except that a coded and punched index card would be produced in lieu of a printed translation. This would eliminate all manual coding and key punching operations with resultant savings in clerical costs and increased speed and accuracy.

The machine would be utilized for decoding the index cards machine searched and selected from an index file to fill a request for document references. In this case the operation would be the opposite of the coding operation described above in that the punched-hole codes contained in the selected cards would be "read", located in the stored "code dictionary" and read out to a typewriter or printer which would type or print the equivalent English word for each code. As a result, the decoded material would serve to assist the requestor in determining which of the referenced documents would be most useful for his needs. It is possible that the results obtained in this manner described might preclude the need for and save the cost of making and maintaining abstracts of documents on a current basis.

5. The above application is only one example of utilizing the operating characteristics of a translation machine for other than language conversion requirements. The ability of the machine to store and process data in accordance with programmed instructions provides for the development of various applications.

D. Recommendations

It is recommended that the Agency support and actively participate in the research and development work being carried on by Dr. Postart.

cc: ~~AND/A~~
Comptroller
OTC

~~CONFIDENTIAL~~

25X1A9a